

餐饮生态大数据架构实践

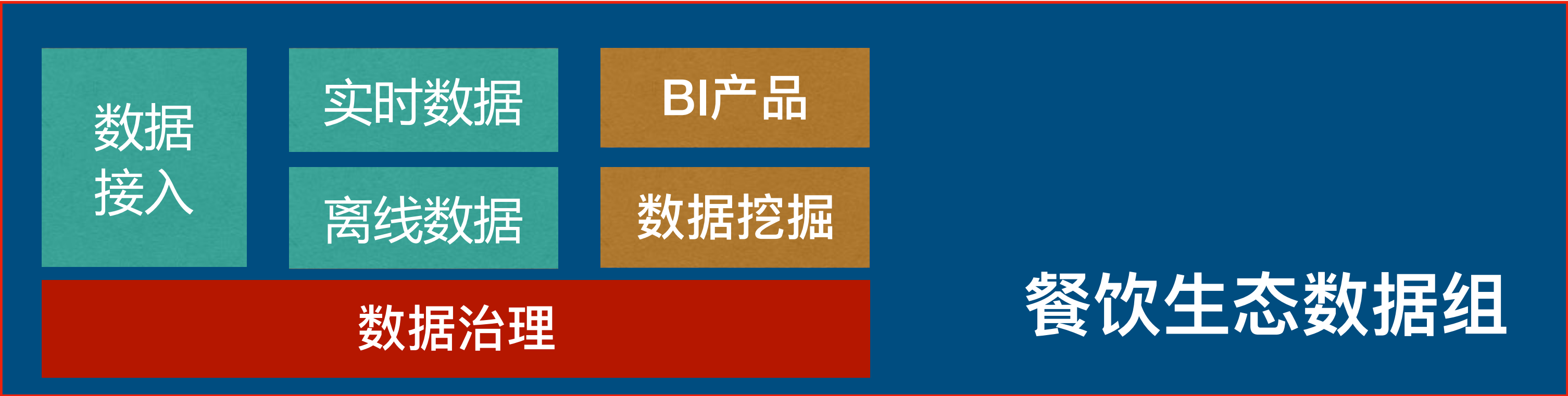
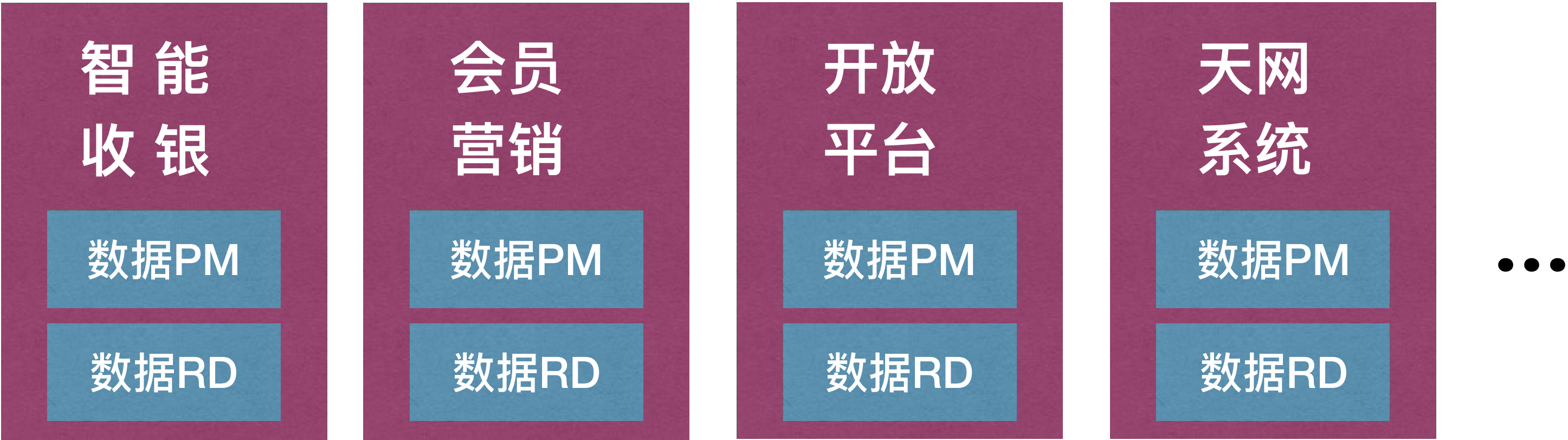
牛江浩@美团技术沙龙

自我介绍

1. 2011~2012 阿里巴巴后台研发工程师
2. 2013 加入原美团
3. 2017 餐饮生态数据应用方向技术负责人

关注点：用数据技术解决业务问题、最大限度的发挥数据价值、支撑并驱动业务发展

团队组织架构



目录

- 离线数据
- 实时数据
- 数据应用
- 总结及思考

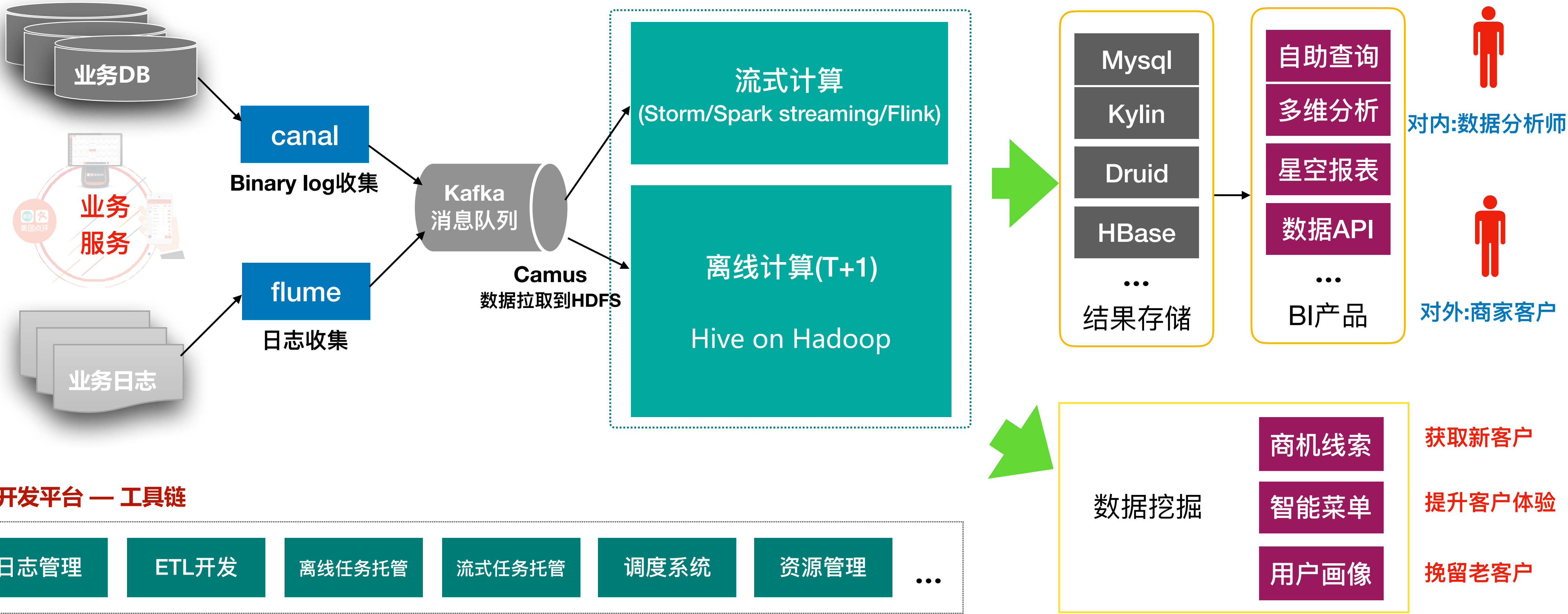


数据流架构全景图

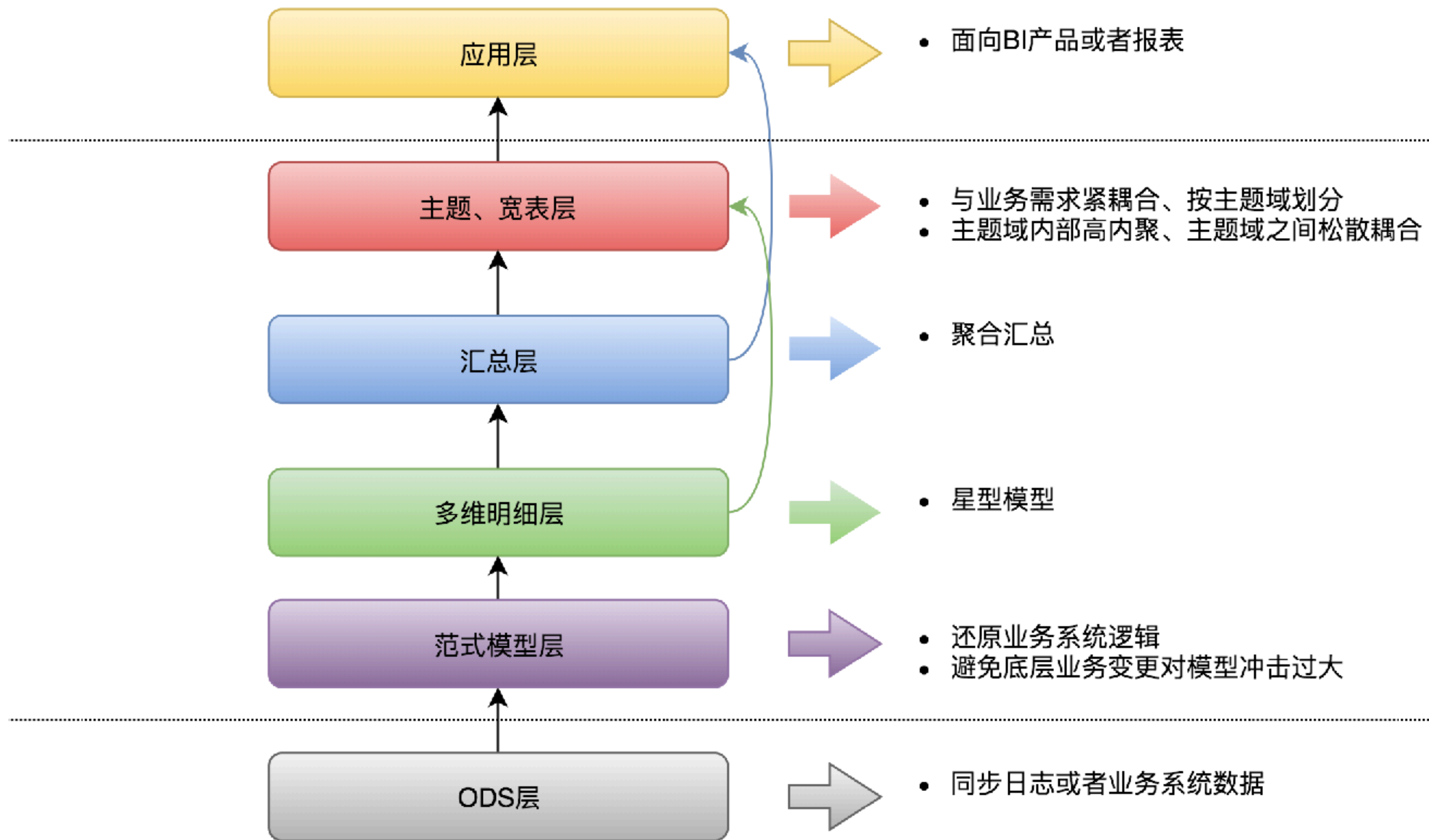
数据接入

数据加工

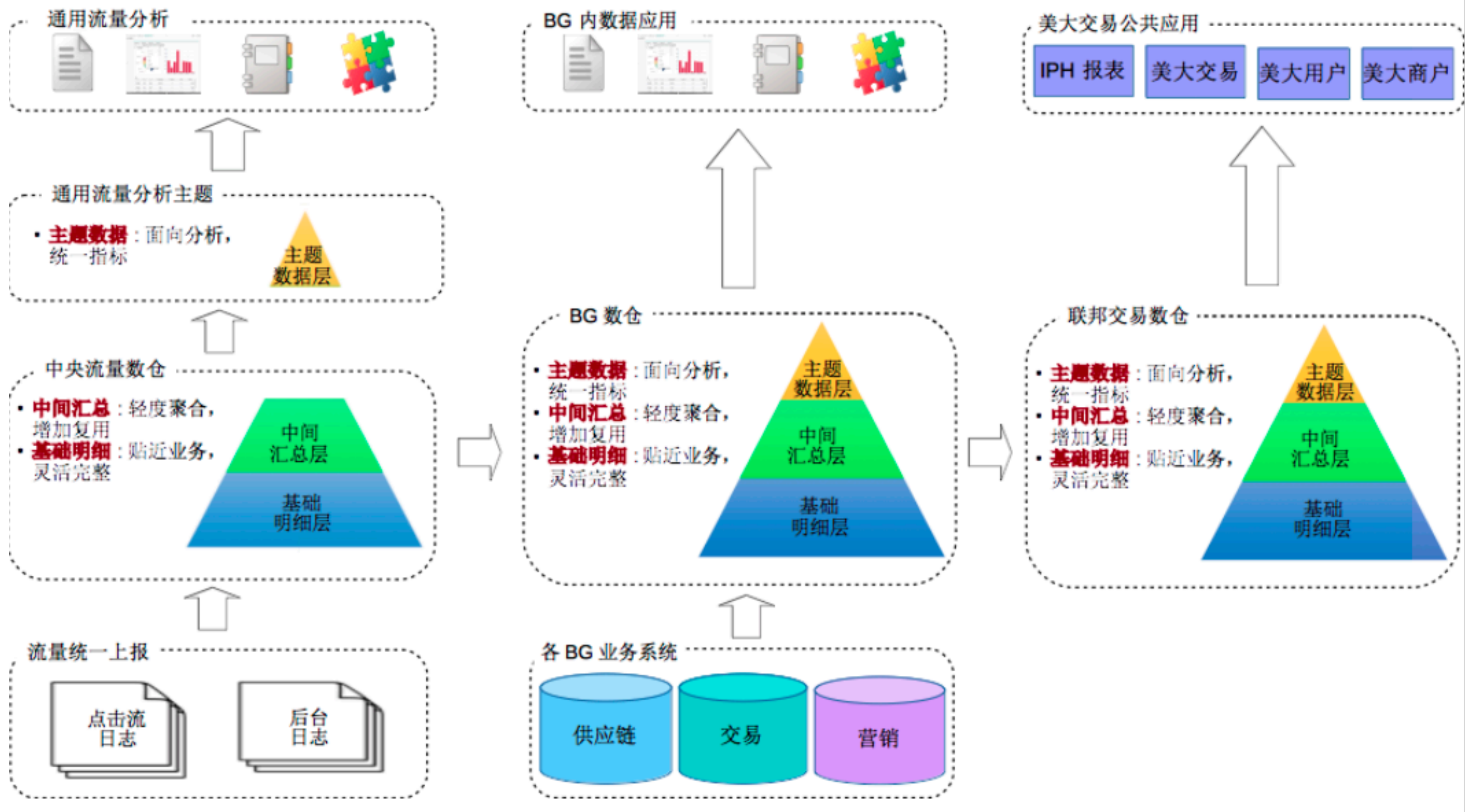
数据应用



离线数据仓库建设



全域数据联动



数据管理

自身业务实际面临的**核心问题**：质量差、效率低、价值少

数据质量管理

生命周期管理

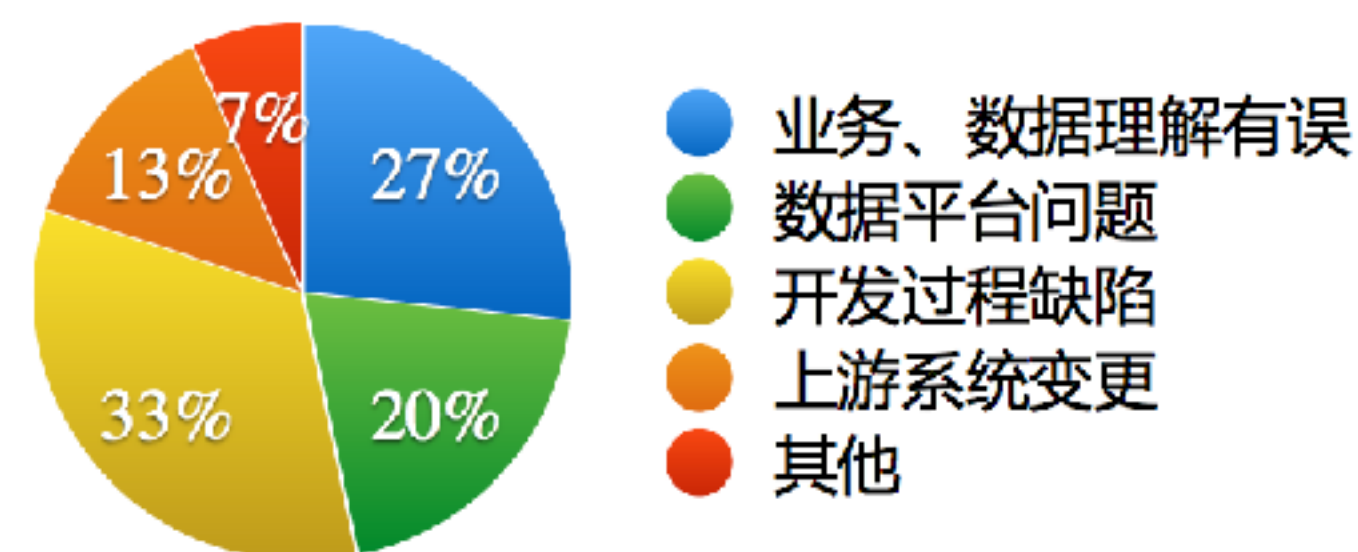
辅助开发工具

数据价值运营

数据管理

数据质量管理-事前识别、预防&规避风险

- **技能培训、加强review、充分测试等** —> 保证开发质量
- **数据设计评审** —> 业务系统、数据、需求方**认知保持一致**
- **确立问责机制** —> 和上游业务系统做好约定、周知义务并**关联KPI**
- 其他诸如和数据平台做好约定、参数变更要周知.....



数据管理

数据质量管理-事中监控、容错、降级

- 离线数据质量监控 —> **及时**发现问题
- **实时**数据质量监控 —> **提前**发现上游问题、提前介入、**避免**问题发生
- **数据质量问题截断** —> 挂起全部、更细粒度的控制 (血缘、影响指标)
- 关键数据进行 **降级、容错**
-

数据内容&时效性监控

- 注册SLA保障
- 数据内容监控

监控设置 hmart_mobile_test.aggr_consume_hot_cate_city_area

表名:	aggr_consume_hot_cate_city_area 🔗字段信息	META:	hmart_mobile_test
ETL名称:		Owner:	--
分区键:		列数:	0

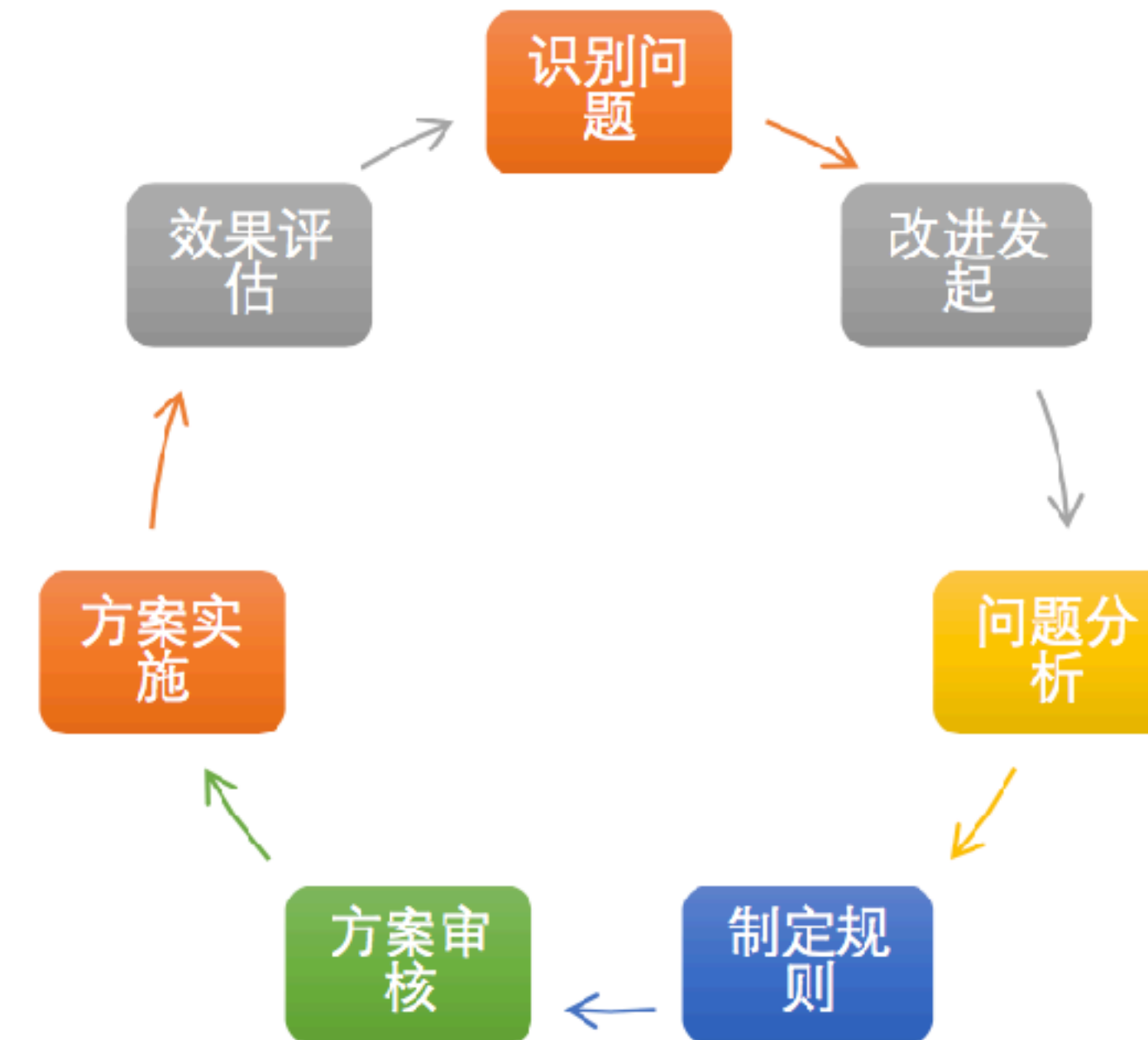
基础配置 [🔗](#)

时间字段:		时间格式:		监控分组:	默认分组
报警接收人:					

数据管理

数据质量管理-事后复盘

- 分析问题**本质**-5why
- 建立反馈**闭环**-PDCA
- 归纳总结**沉淀**-规范



识别、定义、解决问题、**系统化解决问题**、验收复盘、持续提升数据质量

数据管理

工具化辅助、自动化提升效率

痛点：规范 执行成本高、效果差



建模工具化

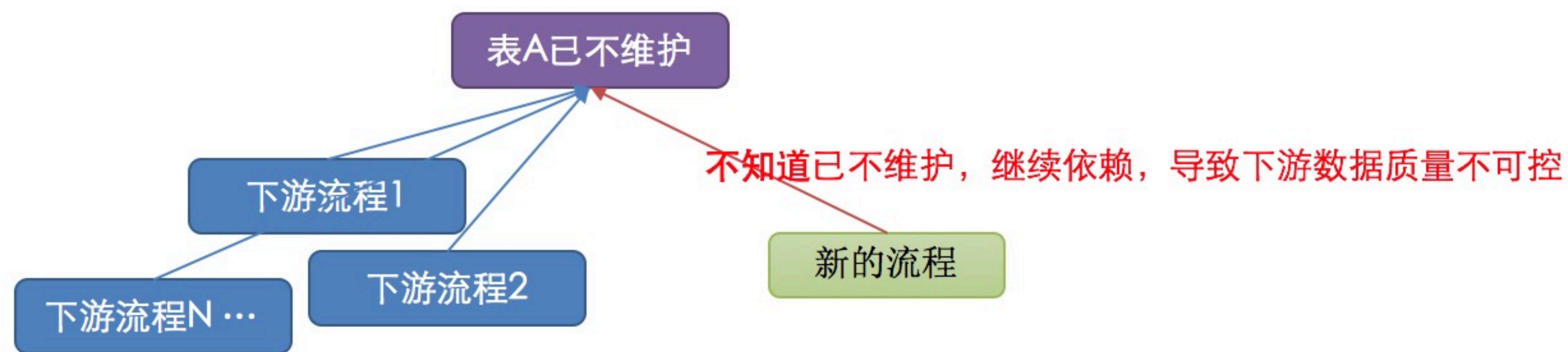
审核工具化

测试工具化

...

数据管理

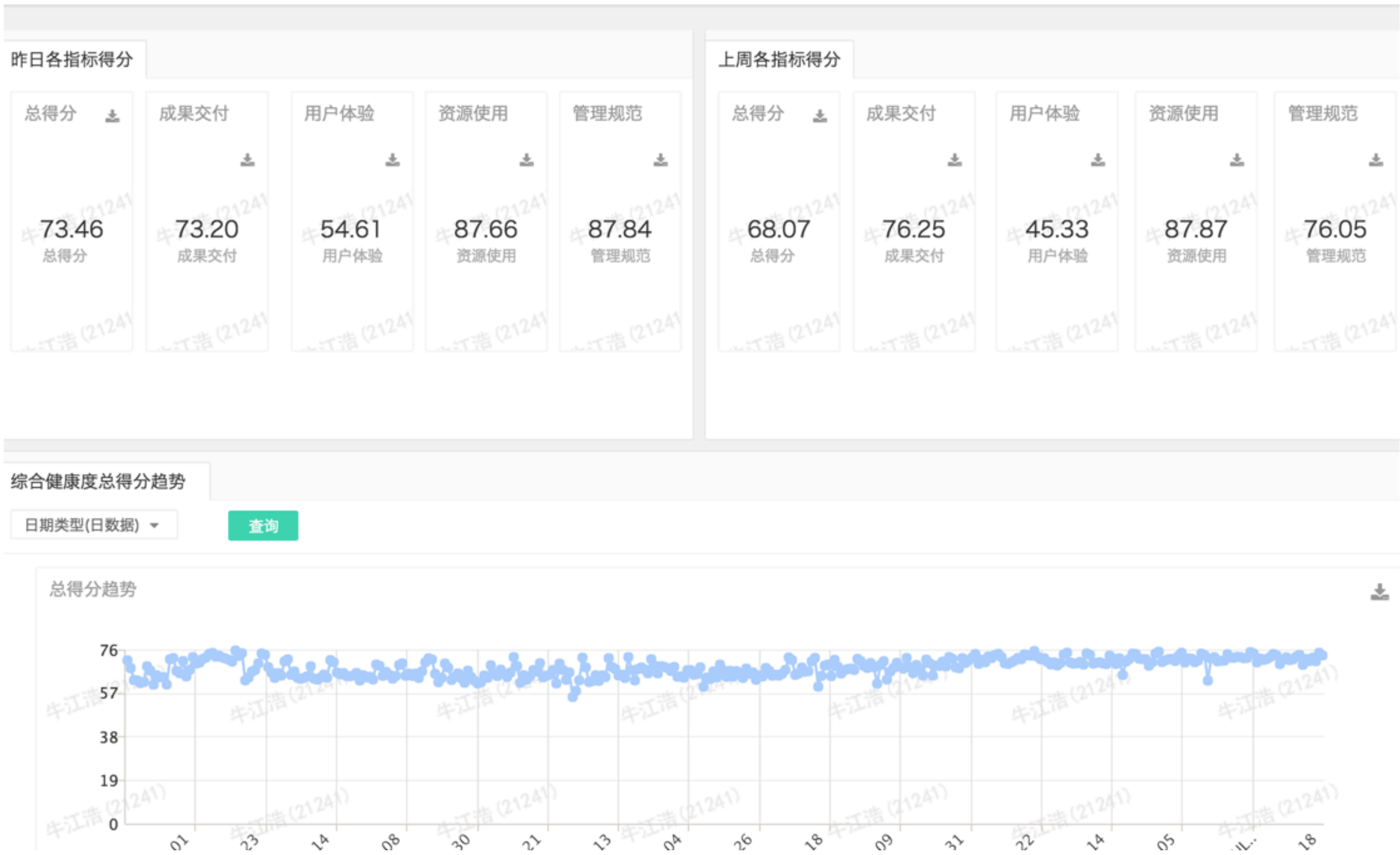
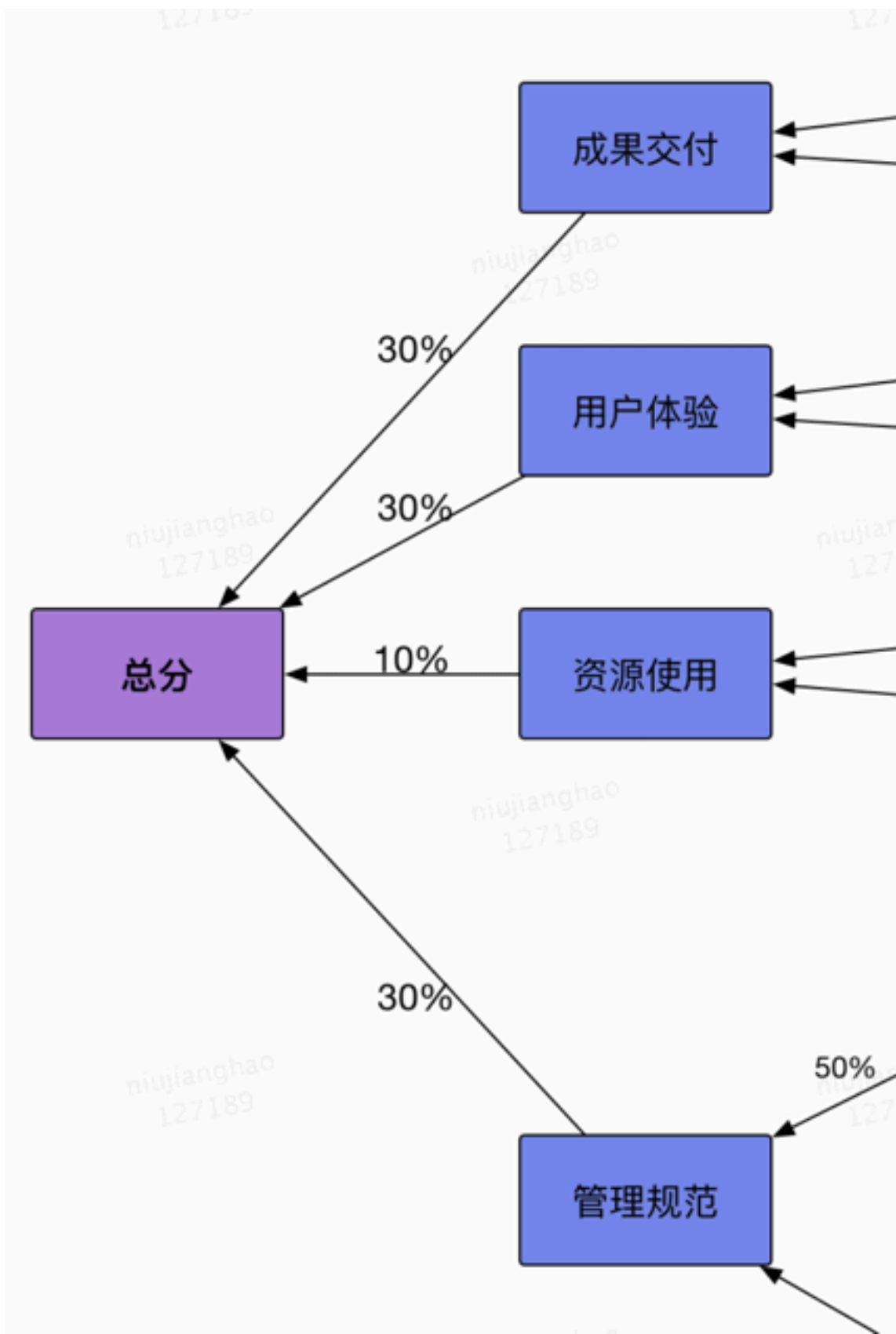
数据生命周期管理



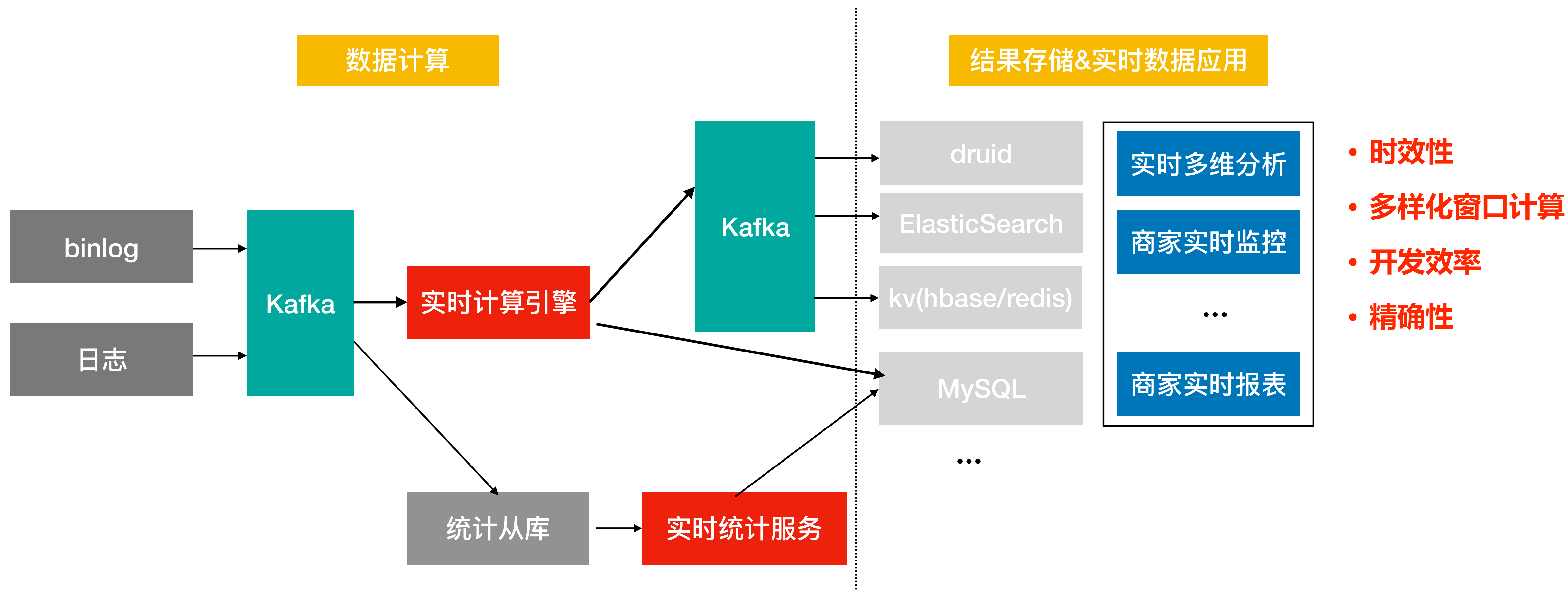
应用元数据 + 数据血缘，及时下线老化、无效数据，保证数据仓库数据整体干净有效

量化数据仓库健康度

通过过程指标为数据仓库建设提供管理和优化的手段

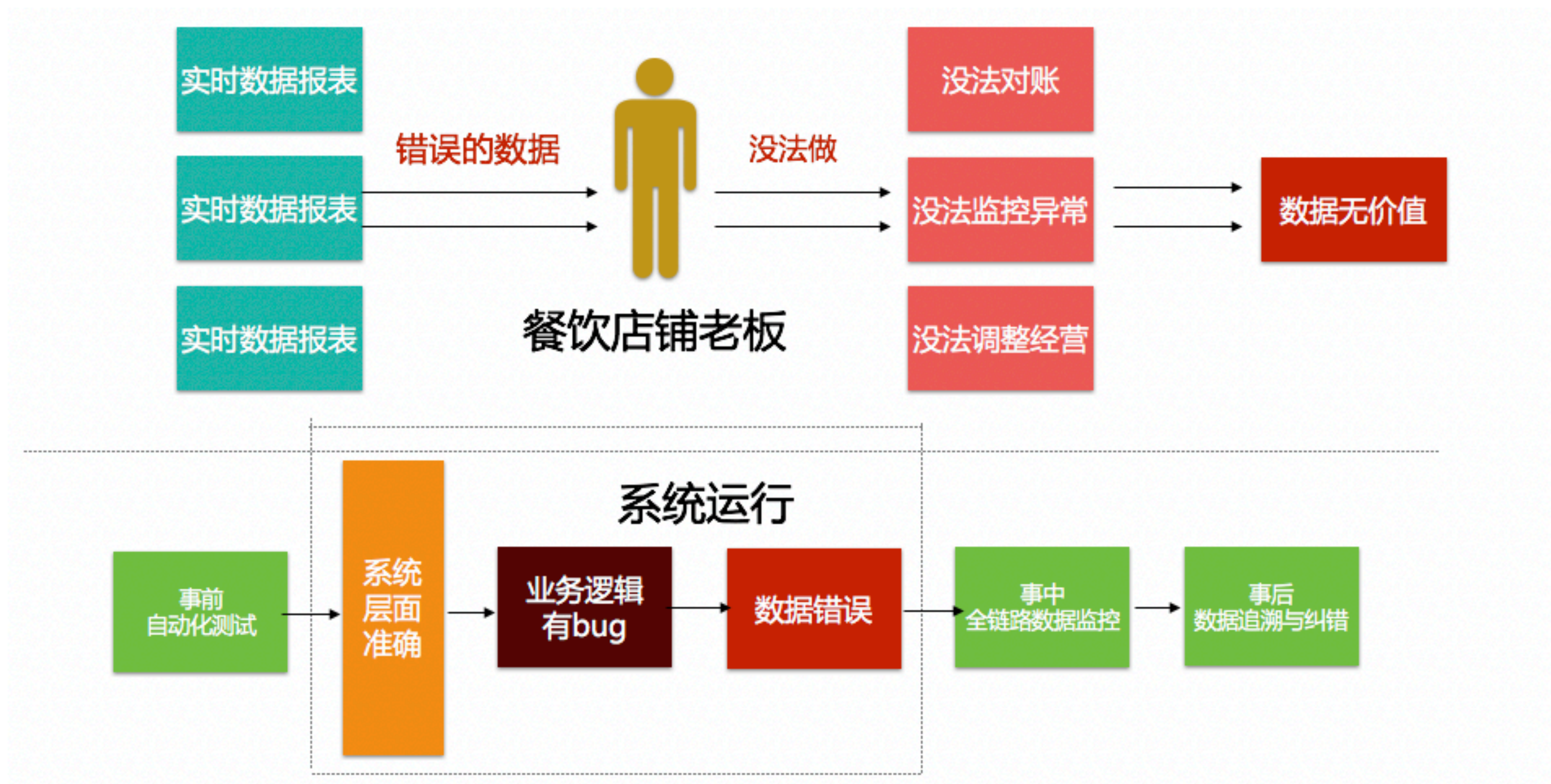


实时数据清洗



实时数据清洗

极高的**准确性**、**可用性**要求



提前识别&规避风险 + 缩短故障发现时间 + 缩短故障处理时间

数据应用

满足内部分析师查看和分析数据的需求

自助查询

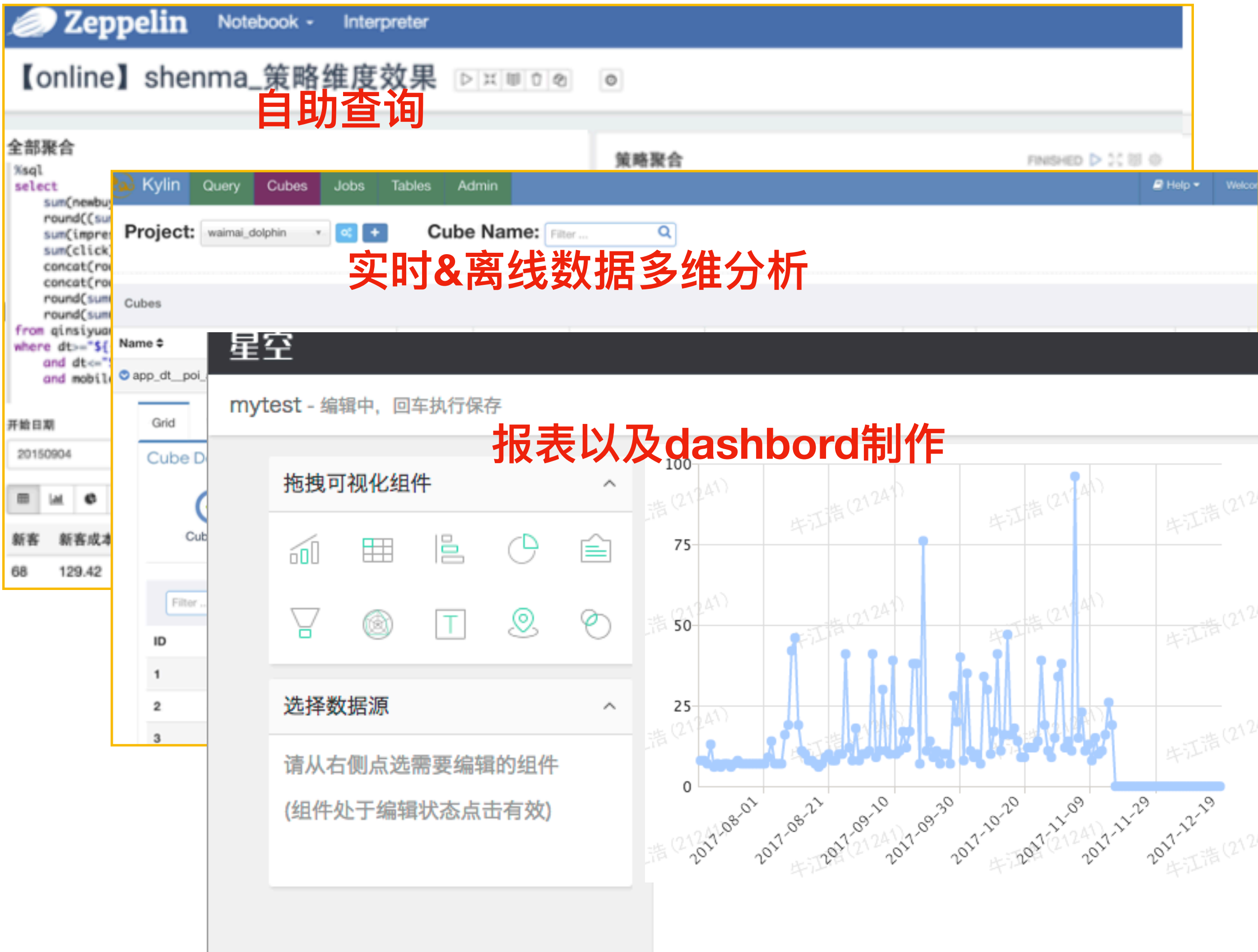
多维分析

星空报表

数据API

...

BI产品



数据应用

商家数据可视化：满足商家经营的需求

自助查询

多维分析

星空报表

数据API

...

BI产品

数据API开发平台



数据API面临的挑战

可用性

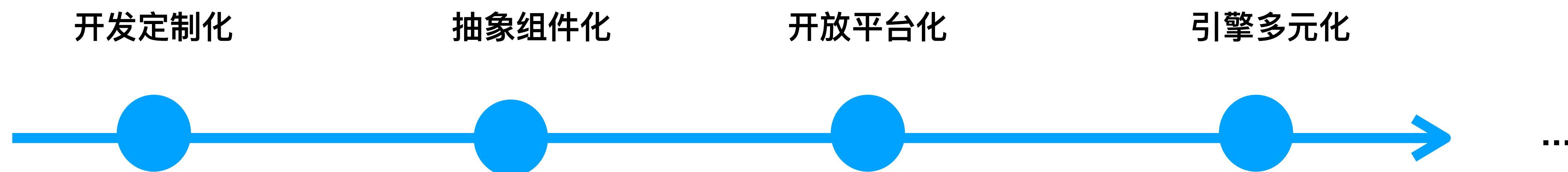
性能

效率

成本

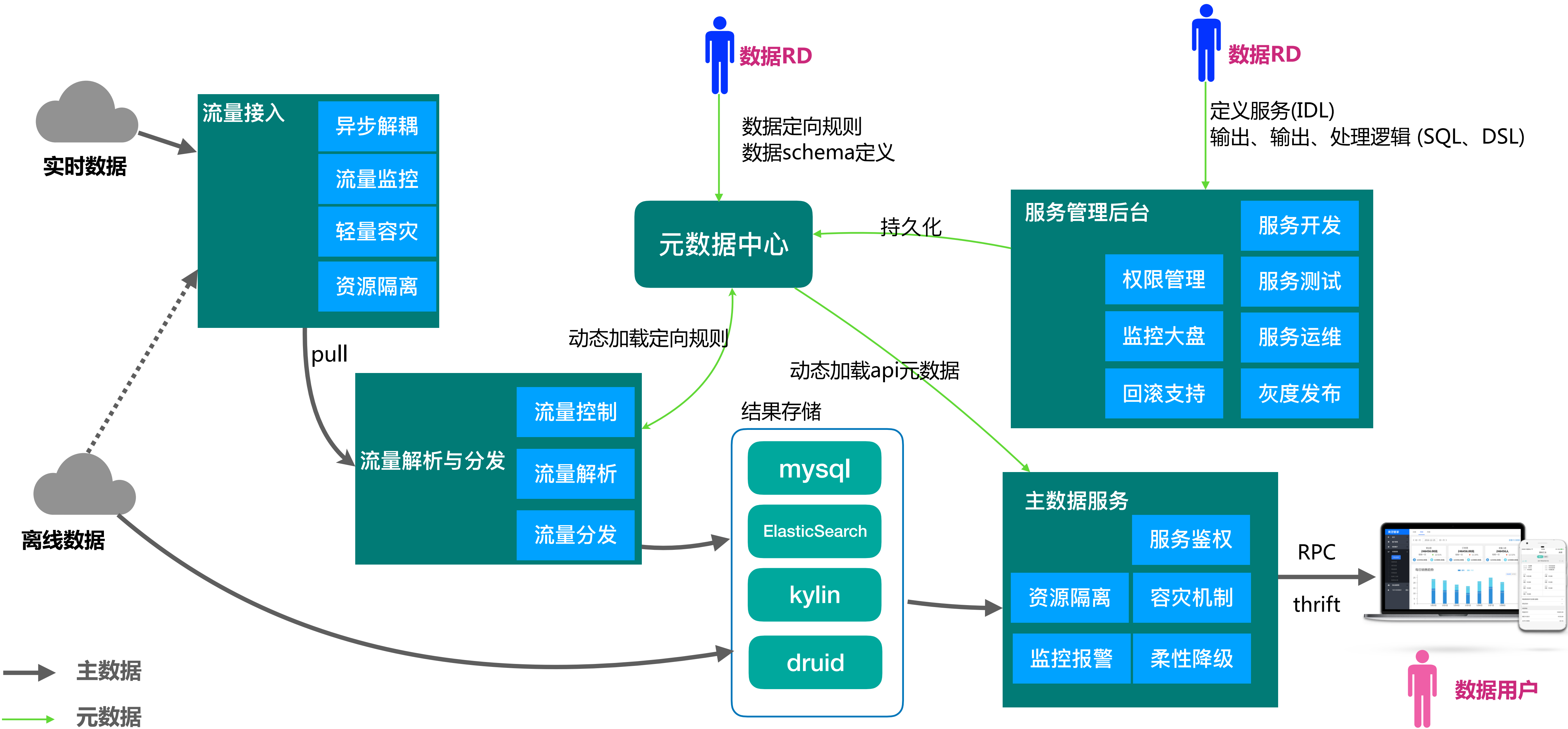
...

数据API开发模式的演进



随着业务的发展，**主动识别&预见** 技术体系中的**痛点**，**下沉抽象解决主要矛盾**

系统逻辑架构



取得效果

- 让不具备后台开发能力的数仓同学具备后台服务输出能力
- 2人力、98%+数据API、累积600+个
- 过去1个Q 接口累积迭代1000余次

总结与思考

- 站在业务&用户的角度看问题
- 面对问题、解决问题
- 升维&降维
- 视野是技术的全局思考力

谢谢大家

Q&A